

Georgia Institute of Technology School of Biology

Finding Evidence of a Cancer Suppressing Mutation, rs 3814113

April 14, 2010

James Small
Undergraduate Student
School of Biology

John McDonald
Professor
School of Biology

Mirjana Brockett
Professor
School of Biology

Introduction:

Genetic mutations are important risk factors for ovarian cancer; with certain mutations increasing a woman's risk for ovarian cancer by as much as 50%. The search for these mutations has recently been led by large projects utilizing the genetic information of thousands of individuals. These studies have been able to find suspected oncogenic mutations such as the Single Nucleotide Polymorphism(SNP) rs3814113, but have been unable to prove causal relationships due to the inability to incorporate hard to quantify data. Smaller studies however are able to incorporate such data and provide definitive results. This study focused on finding evidence to corroborate the relationship between rs3814113 and ovarian cancer risk. This study investigated the lifestyles, personal history, family history, and medical history as well as the genotype of twenty patients to determine their individual risk for ovarian cancer. This information was used in a series of statistical tests to explore the possibility that each patient's risk for ovarian cancer is affected by the presence of rs3814113. In particular the allelic frequencies were calculated and compared to the general population, T-tests were conducted to determine significant differences between groups, and the presence and risk of ovarian cancer were compared to the presence of rs3814113. None of the statistical tests performed were able to provide evidence of a link between rs3814113 and ovarian cancer risk, thus this study could find no evidence of rs3814113 having an effect on ovarian cancer risk.

Current Understanding of Ovarian Cancer

Ovarian cancer comprises 3% of novel cancer cases in women(Jemal, Siegel et al. 2008). This cancer accounts for 5.7% of mortality among female cancer patients, estimated to be 15,000 women in 2008 alone and has a mean survival time of 5 years(Jemal, Siegel et al. 2008). Due to the lethality of this cancer it is among the most well studied. Mutations commonly found in the

tumors, the effects of the cancer, and the antigens it creates are all foci of research. The fruits of this research, genetic testing, hormone replacement therapy, and cancer antigen(CA) 125 testing have saved thousands of lives and helped reduced ovarian cancer's lethality. Despite these previous efforts, questions still remain regarding ovarian cancer. The most sought after of these questions are the mutations responsible for ovarian cancer.

Known mutations for Ovarian Cancers

The Breast Cancer Associated(*BRCA*) genes are the most important oncogenes found to be associated with ovarian cancer. *BRCA* mutations occur in 40% of familial ovarian cancers (Palma, Ristori et al. 2006) and their presence drastically changes the likelihood of a woman contracting ovarian cancer over the course of her lifetime. A woman without *BRCA* mutations normally has a 1.8% risk of developing ovarian cancer, but if a woman has a severe mutation in her *BRCA* genes then her risk of contracting ovarian cancer increases to 44%(Palma, Ristori et al. 2006; Bordeleau, Panchal et al. 2009). This radical change in risk can be explained through *BRCA*'s suspected function as a component of DNA repair. DNA repair pathways thwart cancer development by preventing cancer causing mutations from accumulating. The *BRCA* genes encode for proteins that are integral to these pathways(Tutt and Ashworth 2002). The *BRCA* genes produce proteins that interact with RAD51, and *BRCA* 1 is assumed to act as a scaffold protein for DNA repair complexes(Palma, Ristori et al. 2006). The exact operation of these proteins is not well understood at this point, but mutations in these genes clearly make their products less effective or completely inoperable. Without the *BRCA* proteins, the DNA repair pathway in the ovaries does not function as effectively and other cancer causing mutations are more likely to develop(Palma, Ristori et al. 2006).

Finding other Cancer Related Mutations

Many scientists have hypothesized that *BRCA* may not be the only significant ovarian cancer causing mutation. As explained earlier, hereditary *BRCA* mutations occur in a significant portion of familial ovarian cancer, but not in all of them. 60% of familial ovarian cancer cases do not involve hereditary *BRCA* mutations, yet these families have recurring ovarian cancer in their history. This simple fact has sparked many investigations to find another significant mutation related to the cancer. While each diagnosis of cancer has its own unique set of mutations(Hainaut 2001), different cancers do have common mutations. Certain SNPs common to many forms of cancer and it has been found that the only SNP's common to ovarian cancer were those involved with breast cancer as well (Quaye, Tyrer et al. 2009). Another study used cDNA microarrays to find similar gene expression across ovarian cancer tumors(Jazaeri 2009). A study focused on the *BRCA* Interacting Protein(*BRIP*) as a possible contributor to ovarian cancer(Song, Ramus et al. 2007). These studies utilized large datasets, but none were able to establish a strong correlation between a particular mutation and ovarian cancer. These studies were able to identify what may be mutations with low genetic penetrance, but lacked the data to prove it. Genetic penetrance refers to the likelihood of a gene to affect the phenotype of the individual with lower penetrance equating to a lower effect that gene has on the individual's phenotype. Often these genes can be present in an individual but may not affect the carrier's phenotype, making a cause and effect relationship difficult to establish. Even *BRCA* gene mutations, which have the highest penetrance of ovarian cancer causing mutations, only bestows a cancer risk of 39%.

Genome Wide Association Studies as a solution

To combat the problem of finding more potentially cancer causing mutations, scientists have turned to utilizing Genome Wide Association Studies(GWAS). These studies function by collecting large sample groups and dividing them by their phenotypic differences. The different phenotypes then have their entire genome sequenced and their genetic variations are recorded. The prevalence of genetic variations among certain groups can then be used to create correlations between a specific phenotype and genetic variations despite low genetic penetration(Manolio 2010). The most recent GWAS to produce significant results was done by this past year and it involved 20,000 participants. This study found a correlation between the C allele of the SNP rs3814113 and cancer suppression(Song, Ramus et al. 2009).

Little information exists concerning the function of rs3814113, but what little information does exist supports the hypothesis that it could affect ovarian cancer risk. Rs3814113 is an SNP found in the basonuclein 2 gene. Scientists believe this gene increases the production of ribosomal RNA (rRNA), as shown by the shared homology between Basonuclein 2 and other known rRNA promoters(Vanhoutteghem, Djian et al. 2004). rRNA is one of the components of ribosomes, the cellular machinery that creates proteins. If basonuclein 2 controls the production of cellular machinery then it could function as a cancer suppressing gene by altering the concentration of proteins; particularly if those proteins are related to cell replication. Basonuclein 2 has already been implicated in the progression of normal esophageal cells into a condition known as “Barret’s Esophagus” and the condition’s further progression into esophageal cancer(Akagi, Ito et al. 2009). Furthermore, basonuclein 2 is only expressed in epidermal cells and sex precursor cells such as the ovaries(Vanhoutteghem, Djian et al. 2004). Given the putative function of the gene, its known association with esophageal cancer, and its expression in only two tissues makes the basonuclein 2 an attractive choice for an ovarian cancer

contributing or suppressing mutation. However; no proof has been discovered that supports a causal relationship between basonuclin 2 or rs3814113, and ovarian cancer risk.

Inconclusiveness of Genome Wide Association Studies

Unfortunately due to the weaknesses of GWAS studies only a proposed correlation exists between rs3814113 and ovarian cancer(Manolio 2010). The main advantage and disadvantage of a GWAS is the volume of information collected. The large amount of data collected can be useful for finding correlations between phenotypes and mutations but it must make thousands of comparisons to reach any results. The volume of comparisons made degrades the power of the test overall and thus less able to prove a connection, requiring that any results be incredibly unlikely to be considered significant. This fact results in GWAS studies being used mostly for finding correlations rather than causations between mutations and phenotypes.

Aside from the large number of comparisons made, GWAS studies also leave out large amounts of information on each individual subject preventing them from proving causation. GWAS studies focus almost solely on a patient's genetic information to reach its conclusions. However; each patient also has a medical history and family history that are significant but often difficult to analyze in bulk. These data are used to screen possible participants or not incorporated into the GWAS. Ramus et al. encountered this problem in their most recent project when they were unable to access detailed histories for all their patients. Without the additional information regarding patient histories the project had difficulty analyzing what other risk factors contributed to the pathology of each patient. Thus despite doing a follow up study regarding BRCA and rs3814113, the study could not find any evidence supporting anything greater than a correlation between the two mutations(Ramus SJ 2010). Because a GWAS study ignores so

much of the data it has the potential to overlook important trends, and weakens the overall findings.

Smaller Studies to Test Connections

While large studies can yield valuable information they can often exclude relevant data that may be important to prove causation. However large studies are useful for establishing correlations that can be further tested by smaller projects. Smaller projects can include information larger studies would have difficulty processing such as detailed family and patient histories. Smaller research projects can also invest more effort per sample and can adapt to unexpected findings in the data. Also, these smaller projects are cheaper and so do not represent as large a potential loss. For all these reasons, smaller projects such as this one are necessary to explore the findings of larger studies.

This project in particular will focus on investigating the findings of a previous GWAS that found a correlation between the SNP rs3814113 and cancer suppression. This study will use difficult to quantify information such as the patient and family histories of the subjects. By including this information this study will be able to determine the expected risk of the patients for ovarian cancer and how the presence of rs3814113 changed those risks. With these risk comparisons this study will be able to determine if the previously found correlation was truly significant.

Materials and Methods:

Selecting Samples and Amplifying DNA

Twenty blood samples were selected from stock at the McDonald cancer research lab. Of these twenty samples ten were from ovarian cancer patients and ten were from patients without

ovarian cancer. The selected samples were the newest available to the lab and accordingly were the least degraded by storage, but the patients were sampled randomly at the time of diagnosis.

Table 1: Heat Cycles used in PCR		
Temperature (°C)	Time (Seconds)	Number of Cycles
95	120	1
95	30	30
52	30	30
72	60	30
72	420	1

Once selected, these samples were amplified through PCR processing to increase the concentration of the SNP rs3814113. The SNP had to be amplified so there would be enough genetic material available for sequencing later in the experiment. The cycles of the

PCR process can be found in table 1,

Table 2: PCR Cocktail Mix		
Ingredient	Amount for one Sample(μl)	Total Amount used(μl)
Distilled H ₂ O	10.54	263.5
Buffer	4	100
MgCl ₂	.96	24
Forward Primer	1	25
Reverse Primer	1	25
dNTPs	1	25
TAQ	.5	12.5

and the recipe for the PCR mixture can be found in table 2. The forward primer used was 5'-

TTTCCTGTTTTCTCCATGC-3' and the reverse primer used was 5'- AGGCTCTAGGGAAAGCCAAG-3'.

The primers used were specifically engineered to bind 86 bases upstream of the target SNP and 299 bases downstream of the target SNP. This was done so the amplified DNA would be large enough to separate using gel electrophoresis but be small enough for quick sequencing.

Purifying and Sequencing the DNA Product

Once the samples were successfully amplified the desired DNA was extracted from the rest of the cellular contents for sequencing. The PCR product was loaded into a 2% agarose gel

for electrophoresis. As the DNA proceeds through the gel the smaller fragments move more quickly while the larger fragments progress more slowly causing the formation of bands. These bands each contain DNA fragments of a particular size. The gel was allowed to run for two hours at which point the target band of 386 bases became distinct. Afterwards the bands were cut from the gel using a razorblade. Once excised the gel band was dissolved according to the directions found in the QIAquick PCR purification kit and the DNA was purified in an elution column. The purified DNA solution was then loaded into labeled vials and shipped to Operon Sequencing Incorporated for sequencing.

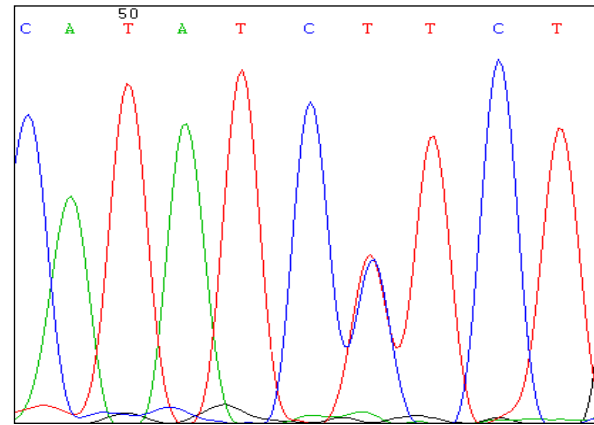


Figure 1: The base signals of a heterozygous patient for the SNP rs3814113

Statistical Analysis

Once the sequenced results were returned each sample was evaluated to determine its genotype. The frequencies of these genotypes were then recorded. Population frequencies of these genotypes were then collected from the International HapMap project, and used to calculate the likelihood of the genotypic concentrations among the pathologic groups.

Patients were split into groups based upon first their pathology and then their genotype to perform all statistical tests. T-tests were also performed to determine if there were any differences between the last menstrual period(LMP) and the age of diagnosis(AOD) among the

genotypic groups and the pathologic groups. Afterwards the prior medical and familial histories of the subjects were collected and used to calculate their individual risk of ovarian cancer using the online program from the Siteman Cancer Center. Race, age, prior cancer history, Familial cancer history, smoking, and history of hormone replacement therapies were used as criteria in calculating the patient's individual risk rating. These risk ratings were used to see if all genotypic and pathologic groups had similar ratios of people with below average, average, above average cancer risk.

Results

The 20 patients and their sequences were grouped initially by their pathology into two groups: those who had developed ovarian cancer and those who had not developed it. Each group then had their genotypic frequencies and the likelihood of these frequencies calculated using population frequencies provided by the International HapMap project. The healthy patients were not found to be statistically different from the general population in their genotypic frequencies, with a heterozygous frequency of 30% and no individuals homozygous for the C allele of the SNP. The Cancer group however differed significantly from the expected genotypic ratios with all but two subjects being heterozygous for the C allele. The resulting binomial distribution resulted in a p-value of .0317, thus showing this group had significantly more heterozygotes than expected and the $P < .05$ level. The allelic frequencies also had interesting result. The control group had significantly fewer C alleles than expected with a p-value of .0288. The cancer groups allelic frequencies did not significantly differ from the population's frequencies.

Next the age and diagnosis for ovarian cancer and the age at last menstrual period(LMP) were evaluated between the two groups. Both of these tests separated subjects into two groups of

ten based on their pathology. In each case the two sample t-tests used assumed unequal variance between the groups, and were two tailed to find differences between the two groups regardless of what those differences were. Neither test found a statistically significant difference between the

Table 3: t-tests and results	
t-test	p-value
LMP by Genotype	0.491
AOD by Genotype	0.333
LMP by Pathology	0.927
AOD by Pathology	0.675

two groups at the p-value = .05 level.

Next the twenty patients were divided into group based on their genotype, this division resulted in one group of 9 homozygous individuals and one group of 11 heterozygous individuals. The t-tests described above were then repeated using these new groups. Again neither test found any significant difference between the two groups.

All results from t-tests can be found in table 3.

Finally the women were evaluated on their individual cancer risk as calculated by the Siteman Cancer Center. These cancer risks were then used to create a table of counts that counted the number of individuals with below average, average, and above average cancer risk in each genotypic group and each pathologic group. All groups were found to have roughly a 4:6:9 ratio. The results can be found in table 4.

Table 4: Cancer Risk Counts.					
Risk	TT	CT	Cancer	Control	Total
Below Average	3	1	2	2	4
Average	2	5	4	3	6
Above Average	4	5	4	5	9

Discussion and Conclusion

Previous studies by Song et al. have supported the correlation between the C allele of rs3814113 and reduced cancer risk, this study however did not find any evidence to support

those findings. When the genotypic frequencies of the two groups were compared to the population averages from the International Hap Map project the cancer group was found to significantly diverge from the population averages. The cancer group showed a much higher prevalence of heterozygotes than expected, with only a 3.17% chance of such a result occurring. The expected result had been that the control group would have a higher prevalence of heterozygotes than the cancer group or the general population. However this experiment has displayed the opposite result were the supposedly cancer suppressing mutation has been found to be more prevalent among the cancer group. The allelic frequencies were likewise surprising in that the C allele was found to be in significantly lower concentration among the control group than expected.

The t-tests showed no evidence to support the C allele of rs3814113 as a cancer suppressing mutation. The T-test for age of diagnosis showed no significant difference between the cancer and control groups. This result indicates that there was no significant difference in onset of the disease and suspected onset of the disease between the two pathologic groups. The second T-test with subjects separated by genotype also resulted in no significant difference between the groups. Thus, there is no significant difference between the onset of ovarian cancer for the two genotypes. When these groups were used to compare the age at last menstrual period there was likewise no significant difference between the groups. Age of diagnosis and LMP are useful tools used to determine if there are important biological differences at work within different patients or groups of people. For instance had there been a difference in the age of onset between the genotypic groups then it could be argued that the presence of the mutations either caused or prevented the onset of cancer. Likewise a difference in LMP could have suggested a change in the gynecological function between the two groups due to the presence of the

mutation. But in this experiment there was no significant difference in LMP regardless of whether the 20 women were divided by pathology or genotype.

Finally the cancer risk of each patient was evaluated using the cancer risk calculator at the Siteman Cancer Center's website. The prevalence of each risk category was calculated for four different groups, the TT genotype, the CT genotype, the Cancer group, and the control group. These ratios were then compared against each to determine if there was a significant difference. Interestingly the ratios across all four groups matched the ratios for the entire group. This result means that all groups had the same ratio of high to low risk cancer subjects, regardless of pathology or genotype and they matched the ratios of the group as a whole. This constant ratio means that currently the presence of the genotype or the pathology of the patients has any effect on the expected ratio of cancer risk.

Between the lack of statistically significant differences for age of diagnosis or LMP between the pathologic or genotypic groups, the uniform cancer risk ratios across the groups, and the higher than expected prevalence of the CT genotype in the cancer group this study can say that it did not find any evidence supporting the C allele as a cancer suppressing mutation.

This study did have some weaknesses. First, this study was very small. While a small study can invest more hours work into investigating each sample than a larger study, these small numbers are susceptible to fluctuations in the data. For example, there was a 3.17% chance that of the ten cancer patients that 8 or more would be heterozygous for the cancer suppressing allele. While this likelihood is small there is a possibility this result is in fact a type I error, and that the sample group collected was not representative of the actual groups of interest. Likewise, the other results could also be affected by fluctuations in the data or the sample group. If this experiment were repeated again it should be done on a larger scale. Thirty or more cancer

patients and thirty women for a control group would provide a large enough sample where groups could be more safely assumed to match the larger population. Larger sample sizes also allow for the use of more powerful statistical tests and greater power in the tests used. A large sample group also gives greater variability to the sample groups that may allow other factors to be tested, for example this study could not evaluate the severity of the cancer across genotypic groups because eight of the ten cancer patients were heterozygous for the mutation.

References:

Akagi, T., T. Ito, et al. (2009). "Chromosomal abnormalities and novel disease-related regions in progression from Barrett's esophagus to esophageal adenocarcinoma." International Journal of Cancer **125**(10): 2349-2359.

Bordeleau, L., S. Panchal, et al. (2009). "Prognosis of BRCA-associated breast cancer: a summary of evidence." Breast Cancer Research and Treatment **119**(1): 13-24.

Hainaut, M. O. a. P. (2001). "**TP53 mutation patterns in breast cancers: searching for clues of environmental carcinogenesis.**" seminars in Cancer Biology **11**: 353-260.

Jazaeri, A. A. (2009). "Molecular profiles of hereditary epithelial ovarian cancers and their implications for the biology of this disease." Molecular Oncology **3**(2): 151-156.

Manolio, T. A. (2010). "Genomewide association studies and assessment of the risk of disease." The New England Journal Of Medicine **363**(2): 166-176.

Palma, M., E. Ristori, et al. (2006). "BRCA1 and BRCA2: The genetic testing and the current management options for mutation carriers." Critical Reviews in Oncology/Hematology **57**(1): 1-23.

Quaye, L., J. Tyrer, et al. (2009). "Association between Common Germline Genetic Variation in 94 Candidate Genes or Regions and Risks of Invasive Epithelial Ovarian Cancer." PLoS ONE **4**(6): e5983.

Ramus SJ, K. C., Gayther SA, Pharoah PD, Sinilnikova OM, et al. (2010). "Genetic Variation at 9p22.2 and Ovarian Cancer Risk for BRCA1 and BRCA2 Mutation Carriers." Journal of the National Cancer Institute **103**: 105-116.

Song, H., S. J. Ramus, et al. (2007). "Tagging Single Nucleotide Polymorphisms in the *BRIP1* Gene and Susceptibility to Breast and Ovarian Cancer." PLoS ONE **2**(3): e268.

Song, H., S. J. Ramus, et al. (2009). "A genome-wide association study identifies a new ovarian cancer susceptibility locus on 9p22.2." Nat Genet **41**(9): 996-1000.

Tutt, A. and A. Ashworth (2002). "The relationship between the roles of BRCA genes in DNA repair and cancer predisposition." Trends in Molecular Medicine **8**(12): 571-576.

Vanhoutteghem, A., P. Djian, et al. (2004). "Basonuclin 2: An Extremely Conserved Homolog of the Zinc Finger Protein Basonuclin." Proceedings of the National Academy of Sciences of the United States of America **101**(10): 3468-3473.